Eliseo Ferrante
Robotics Meeting
May 22th 2008

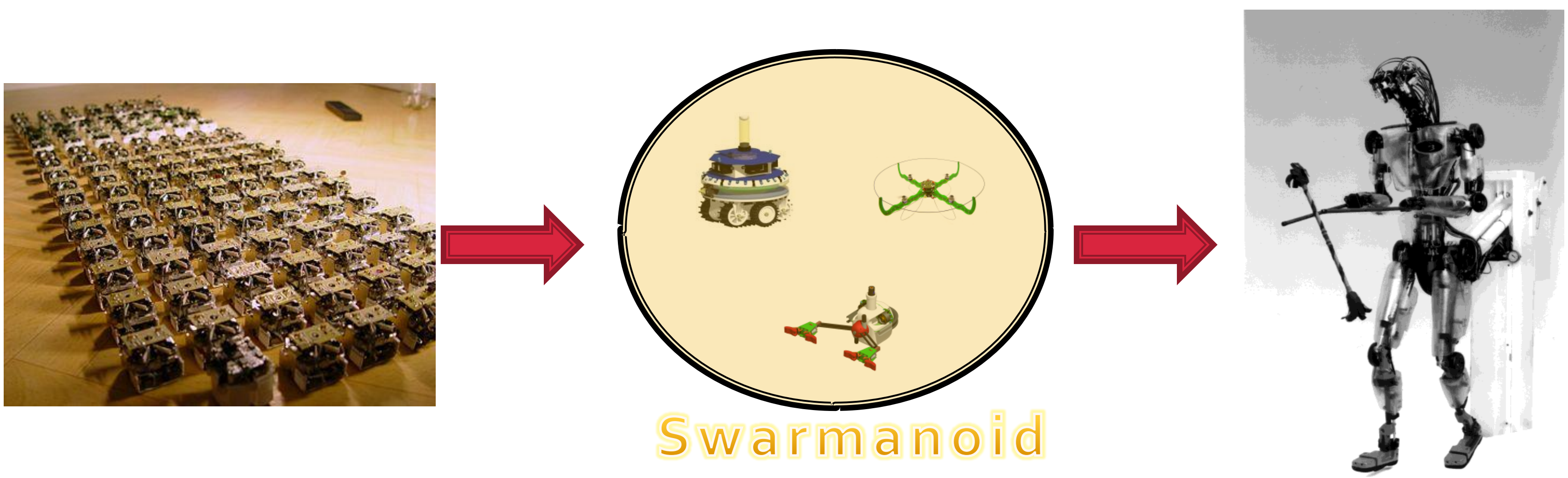
Distributed Perception and Autonomous Learning in Swarmanoid

Outline

- Introduction
- State of the Art
  - Swarmanoid Project
  - Adaptive Routing
- Research Proposal
  - Eye-bots: Distributed Perception and Reward Management
  - Foot-bots: Autonomous and Incremental Learning of Hierarchical Controllers
- Conclusions
- References

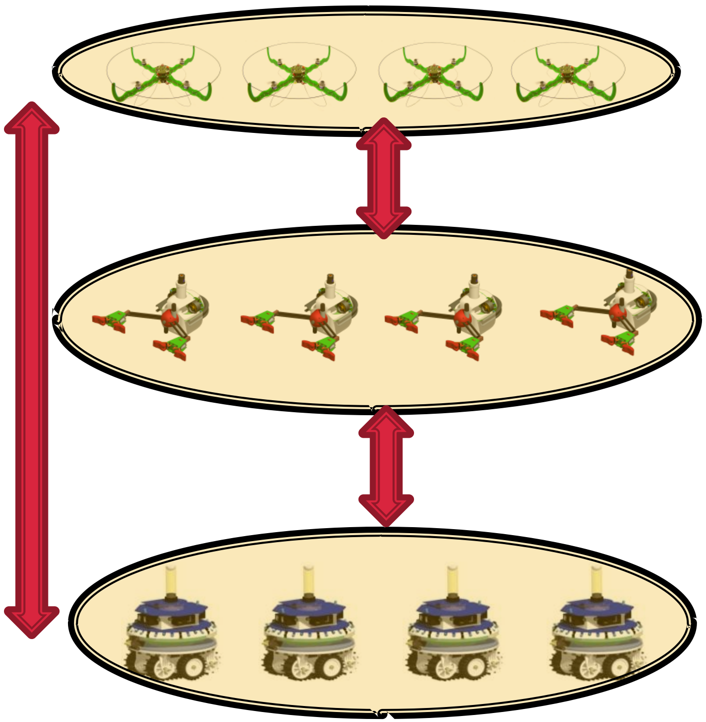
Introduction
The Swarmanoid Project

- Builds on top of swarm-bots
- Towards humanoid swarms through specialization

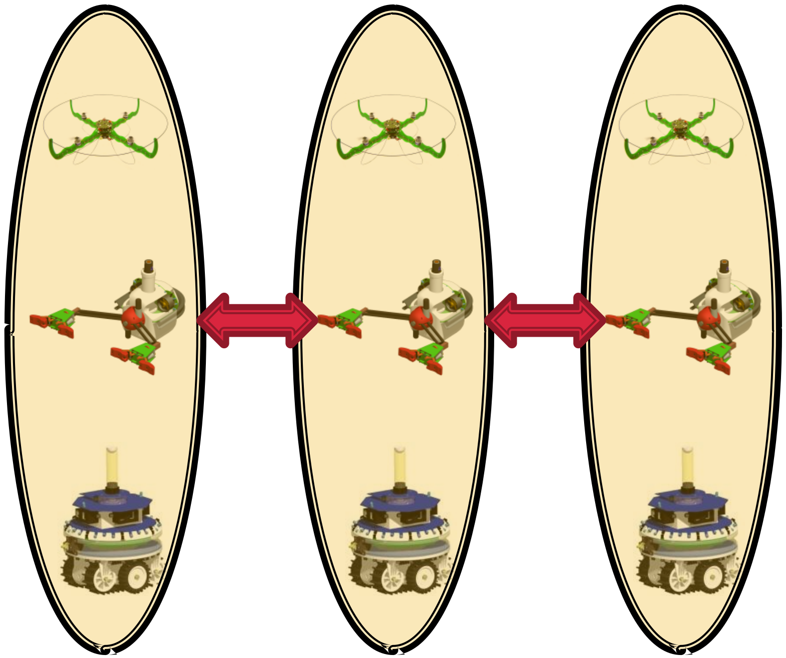


Introduction
Heterogeneous Swarming

HORIZONTAL SWARMING

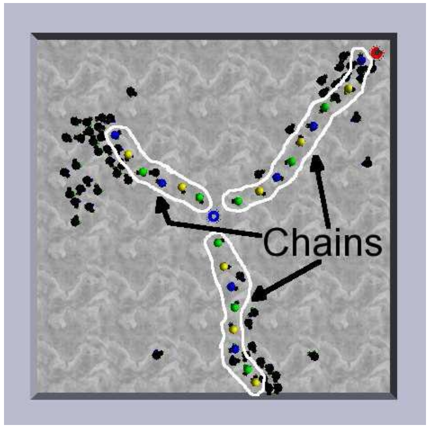


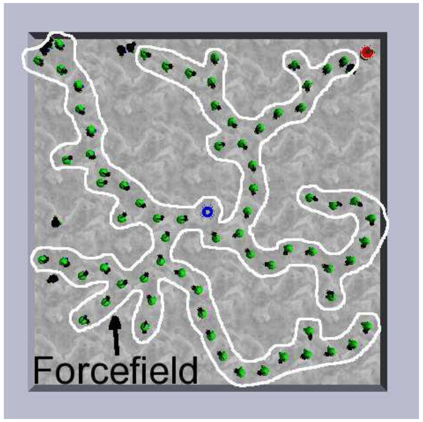
VERTICAL SWARMING



State of the Art - Swarmanoid
Robot chaining

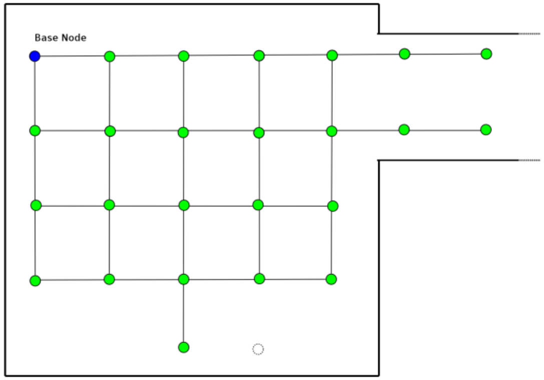
- Developed by Shervin et al. [9] inside Swarm-bots
- Chains (Force fields) are formed between a nest and a prey
- Simple communication schemas (LEDs) are used to provide navigation
- Probabilities of leaving/joining structures are used to ensure continuous exploration
- The final spatial structure “embeds the entire task”

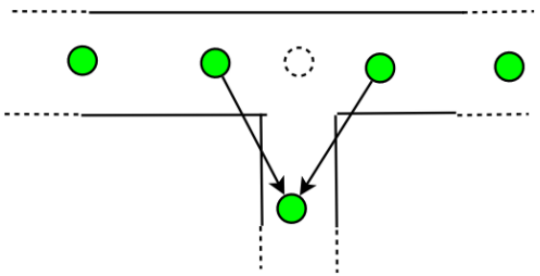




State of the Art – Swarmanoid
Eye-bot's sensor network

- In development for Swarmanoid by Timothy Stirling [10] (EPFL)
- Deployment of a distributed (sensor) network of communicating robots in in-door environments
- Used (mostly) for navigation and inference of world topology
- Result: regular square-lattice formation
- Relies on a presence of a target (task) (?)

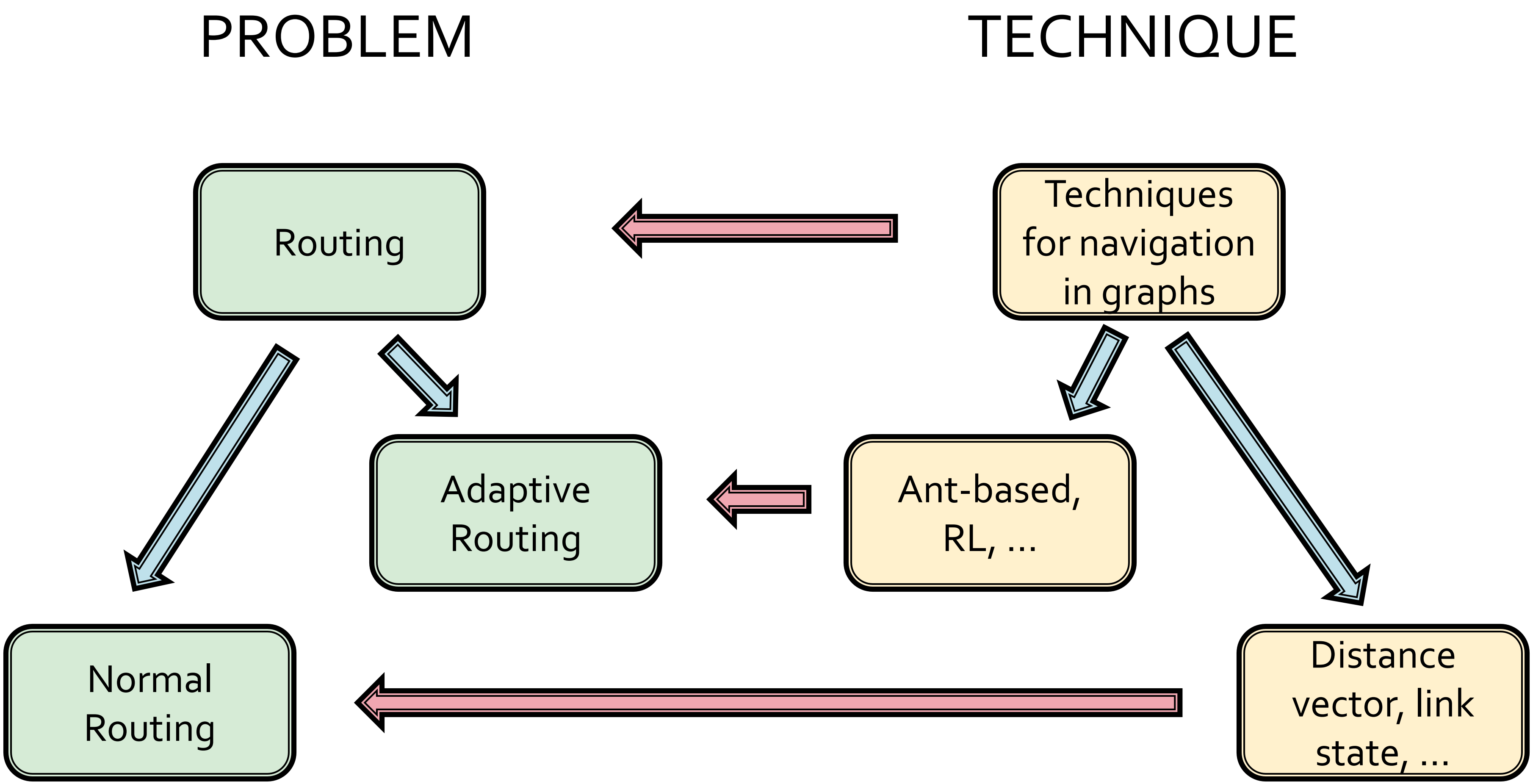




State of the Art – Adaptive Routing
Problem statement

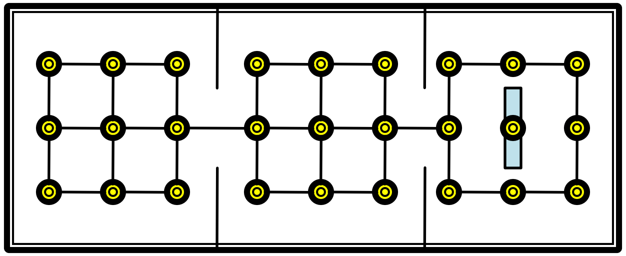
- Routing: find the shortest path in a network (graph) and provide forwarding rules
- "Static" networks can benefit from standard routing protocols
- Adaptive routing outperforms non-adaptive routing in networks with changing topology/connectivity (i.e. MANETs)
- Ant-based and reinforcement learning (RL) techniques have extensively been used to tackle the problem ([1] to [8])
- Idea: reward-based mechanisms produces/learns a routing table in each node
- Used also in robotics for navigation

State of the Art – Adaptive Routing Classification of techniques



Research Proposal
Chaining VS "Task-less" Diffusion

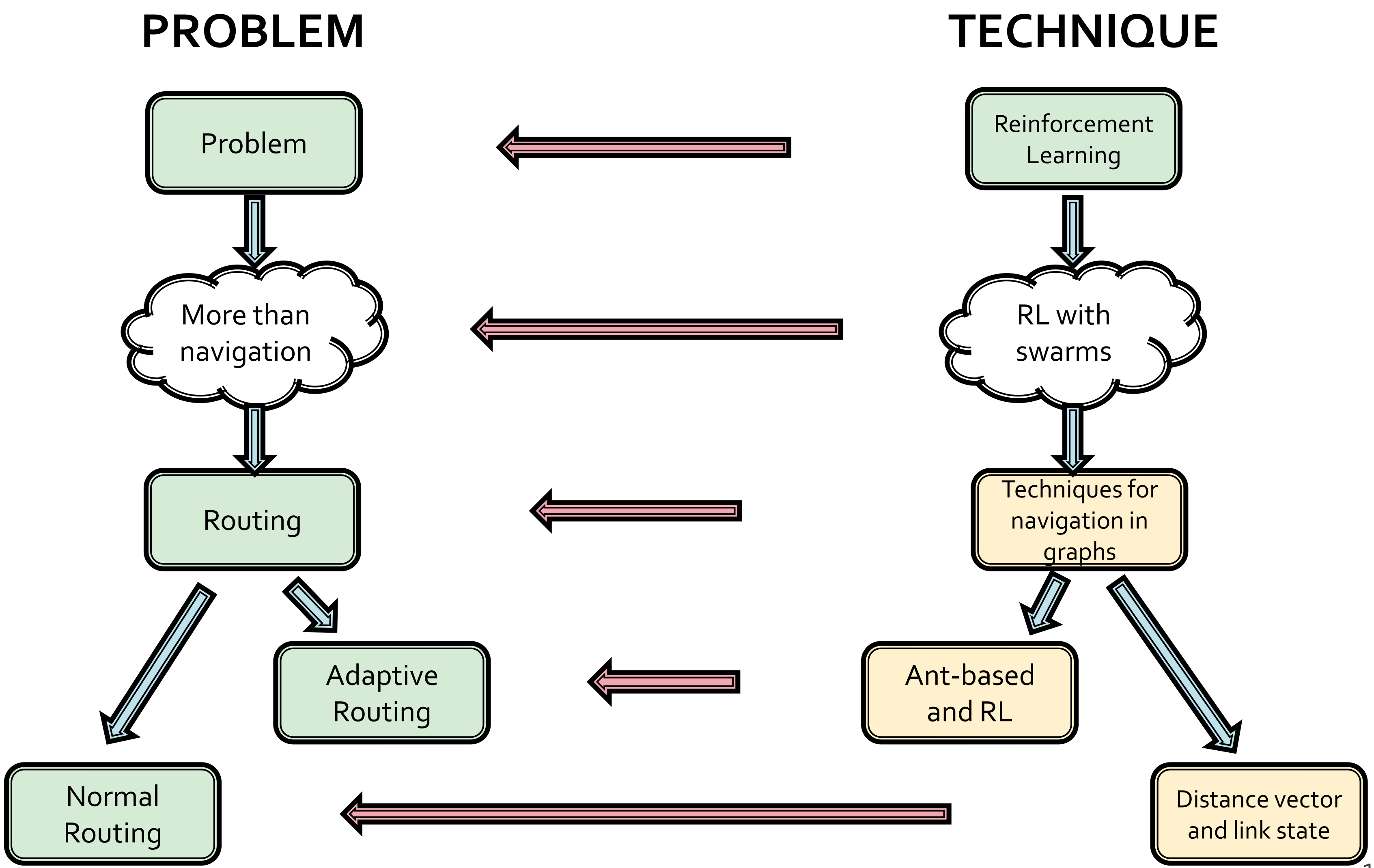
- Builds on top of robot (eye-bot) chaining
- Idea: diffuse a swarm of eye-bots in a (in-door) environment and enable their self-organization into simple-structured lattices that take into account the spatial constraints of the environment
- Does not rely on the presence of a target (task-less)
- Result: distributed graph model = perceived representation of the environment



Research Proposal
Eye-bots as Reward Managers

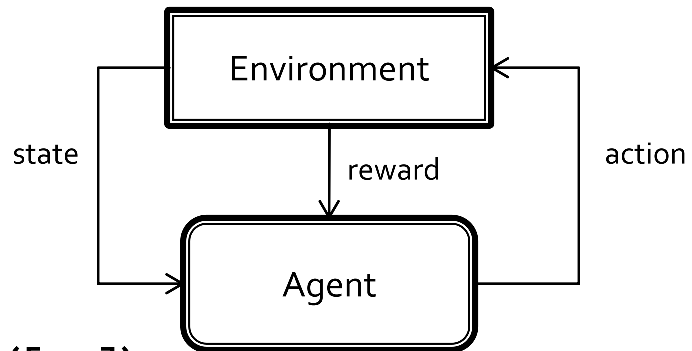
- A swarm of eye-bots can be a support for autonomous learning/evolution/adaptation of foot-bot controllers
- Idea:
  - Use individual eye-bot sensory data to detect features of the environment
  - Based on this data, build a reward or fitness function that express how good it is to be in that eye-bot's area
  - The reward is sent once the foot-bot's presence is detected/signaled
- This idea is not restricted by the particular technique used for learning ...
  - ... but it naturally fits the RL framework
- Can we do more than navigation?

Research Proposal
Classification of Techniques (Revisited)



Research Proposal
Reinforcement Learning

- Learning of an optimal behavior (policy) by direct interaction with the environment [12]



- MDP : <S, A, P, R> ([11])
- The value function $V^{\pi}(s)$ ($Q^{\pi}(s,a)$) contains the long-term value of the state (of the state-action pair) s (s,a)
- The optimal policy $\pi^{*} : S \rightarrow A$ maximizes $V^{\pi}(s)$ or $Q^{\pi}(s,a)$
- RL objective: find $V^{*}(s)$ or $Q^{*}(s,a)$
- Very general framework, but practically used to solve simple tasks (mostly discrete) due to complexity

Research Proposal
Autonomous and Hierarchical Learning

- Eye-bots are used only for ambient shaping and reward management
- Foot-bots autonomously learn controllers through exploration and by simple interaction with eye-bots
- Controllers are then stored and can be available later as macro-actions
  - Using the Markov options framework [13] $o :< I, \pi, \beta >$
- Advantages:
  - In navigation, we can tackle multiple targets (even with different importance)
  - More than navigation
  - Eye-bots do just perception (saving power)
  - Controllers can be hierarchically organized

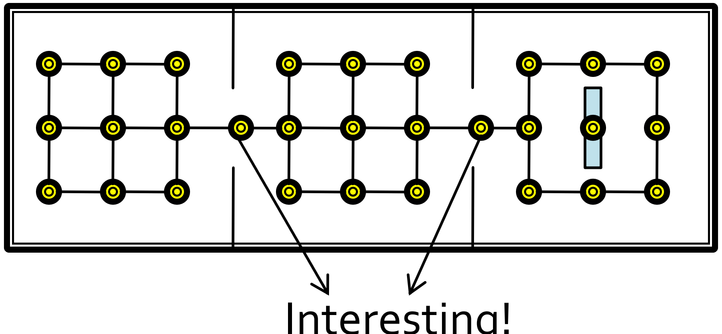
Research Proposal

More than navigation

- We can relax the condition that binds one eye-bot with one state
- Combine with:
  - Internal states
    - Battery level
    - Actuators status (gripper open/closed)
    - …
  - External states
    - Morphogenesis shape
    - Other landscape info
    - …

Research Proposal
Intrinsical Motivation

- Alternative to traditionally hard-engineered injection of reward
- Autonomous agents autonomously and incrementally build skills (i.e. options) by self-assigning rewards through self-motivation ([14] to [18])
- Idea: define an interest function $I(s)$, detect interesting states and assign high rewards to those states
- In Swarmanoid: use eye-bot lattice and graph-based approaches ([16] to [18]), i.e. to detect bottlenecks



Research Proposal
Possible implementation in Swarmanoid

- Deploy the eye-bot sensor network (ceiling attachment)
- Foot-bots navigate the environment
  1. The foot-bot propagates a signal in the environment
  2. The eye-bots who gets the signal answers propagating the state and the reward
  3. The foot-bot considers, as its current state, the one associated with the strongest signal and disregards the others
  4. It selects amongst a set of available actions, implemented in lower-level controllers, which is fixed if the lattice structure is regular (alternatively it can be sent by eye-bots and/or it can include macro-actions)
  5. After a certain amount of time (or other criteria), goto step 1
- Once we defined how <s,a,r,s’> are gathered, theoretically any RL technique can be used for learning

Conclusions

- A set of research directions in swarmanoid:
  - Distributed perception and reward management using swarm of eye-bots
  - Autonomous Learning using RL
    - Alternative to state-of-the-art techniques in swarm robotics
    - Takes advantage of the "discretization" performed the eye-bots swarm
- Somehow big project addition, but decomposable in relatively small self-contained sub-projects:
  - Improve existing work on eye-bots sensor networks
  - Try RL for navigation
  - …

Conclusions

- Somehow big project addition
  - → need to carefully organize it into sub-projects, allocate time, if obstacle then refine initial assumptions/choices/goals, etc…
- What about foot-bot swarming?
  - Take MARL and Social Learning into account
- What about hand-bots?
- What about adaptivity?
  - Use Transfer Learning in RL

References - 1

ADAPTIVE ROUTING

- [1] J. Boyan and M. L. Littman. Packet routing in dynamically changing networks: A reinforcement learning approach.
- [2] Y.H. Chang, T. Ho, L. P. Kaelbling. Mobilized ad-hoc networks: A reinforcement learning approach
- [3] G. Di Caro, F. Ducatelle, and L. Gambardella, "AntHocNet: An ant-based hybrid routing algorithm for mobile adhoc networks,"
- [4] A. Egorova-Forster1 and A. L. Murphy. A Feedback-Enhanced Learning Approach for Routing in WSN
- [5] L. Peshkin and V. Savova. Reinforcement Learning for Adaptive Routing
- [6] P. Wang and T. Wang. Adaptive Routing for Sensor Networks using Reinforcement Learning.
- [7] D. Subramanian, P. Druschel, J. Chen. Ants and Reinforcement Learning: A Case Study in Routing in Dynamic Networks
- [8] Collection of related papers http://citeseer.ist.psu.edu/article/subramanian97ants.html

SWARMANOID

- [9] S. Nouyan and M. Dorigo. Path Formation in a Robot Swarm.
- [10] Swarmanoid Deliverable D10 (WIP) - Not linkable

References - 2

REINFORCEMENT LEARNING

- [11] R. Bellman. A markovian decision process.
- [12] R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction.
- [13] R. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: Learning, planning, and representing knowledge at multiple temporal scales.

REINFORCEMENT LEARNING – INTRINSIC MOTIVATION

- [14] A. G. Barto, S. Singh, and N. Chentanez. Intrinsically motivated learning of hierarchical collections of skills.
- [15] M. Restelli, A. Bonarini, and A. Lazaric. Learning reusable skills through self-motivation.
- [16] I. Menache, S. Mannor, N.Shimkin. Q-Cut - Dynamic discovery of sub-goals in reinforcement learning.
- [17] O. Simsek, A.P. Wolfe, A. G. Barto. Identifying Useful Subgoals in Reinforcement Learning by Local Graph Partitioning
- [18] O. Simsek, A.G. Barto. Using relative novelty to identify useful temporal abstractions in reinforcement learning.

Feedback

Any questions?
Feedback?
Discussion?